

caBIG

Cancer Data Standards Repository (caDSR) IV End-To-End



June 9, 2004 Denise Warzel NCI, Center for Bioinformatics



Agenda

- Introductions
- Goals of Session
- caDSR Overview
- caDSR and EVS End-to-End
- **▶** Transforming UML Models into caDSR Metadata





Introduction/Assumptions

- Denise Warzel Program Manager for caDSR Software Development
 - caDSR is part of caCORE Dr. Peter Covitz, Director, Bioinformatics Core Infrastructure
 - Formerly with IBM for 20 years
 - Background in applications programming and database design, software and process engineering, database migration and services and marketing
 - Started working with NCI on CDE project in 1999
 - Facilitated early CDE development efforts for various lung cancer consortiums
 - Began managing caDSR software development in 2002, starting with the Curation Tool and now responsible for all caDSR repository and tool development activities
- Assumptions
 - Familiarity with CDEs and caDSR Tools





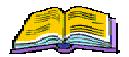
Goals of this session

- Provide a brief introduction to the NCI Cancer Data Standards Repository (caDSR)
 - Describe history and purpose
 - Review the metamodel which forms its basis
- Provide a brief overview of Common Data Elements (CDEs)
 - What is a data element?
 - How is a data element formed?
- Discuss CDEs End-to-End
 - EVS linkage
 - UML Models





Pre-Test → Name the Icon



- Enterprise Vocabulary System
- "The Bible"



- cancer Data Standards Repository
- "The Birthday Cake"



caDSR Admin Tool



caDSR CDE Browser



caDSR Curation Tool







caDSR Background

- Desire to share and browse standardized terms to describe data collected for cancer clinical trials, almost exclusively by and for the Clinical Trails Evaluation Program (CTEP)
- Collections of data elements were "common" within specific diseases, trial types and phases = "common data element" = CDE
- ▶ Original Tool → CDE Browser
- CDEs caught on → more robust set of metadata and tools
- CDE Steering Committee was formed to assess the needs of the growing user community
 - Questions raised:
 - What attributes were needed to describe a piece of data unambiguously?
 - What type of tool(s) is needed?
 - How do we communicate to potential users which CDEs are "good" and which ones are still being developed? Which ones should they have access to and what should the developmental phases be called?







caDSR Background

- CDE Steering Committee was introduced to * ISO/IEC 11179 Information Technology Data Management and Interchange – Metadata Registries 11179 Parts 1-6
 - Roadmap for metadata development
 - ISO 11179 Part-1, Framework and specifications
 - ISO 11179 Part-2, Classification of data elements
 - ISO 11179 Part-3, Basic attributes of standardized data elements
 - ISO 11179 Part-4, Rules and guidelines for formation of definitions
 - ISO 11179 Part-5, Naming and identification principals for data elements
 - ISO 11179 Part-6, Registration of data elements
- CDE Steering Committee voted to adopt ISO/IEC 11179 and transform the CDE Browser and its contents into an ISO Compliant Metadata Repository
 - Benefits included:
 - Accelerate metadata development
 - Simplify ability for users to learn and understand caDSR metadata
 - Enable sharing of metadata with other metadata repositories
- Former caDSR monikers:
 - CDE Browser, DER (Data Element Repository), SBR (Standards Based Repository)

*ISO (the International Organization for Standardization)/IEC (the International Electrotechnical Commission)







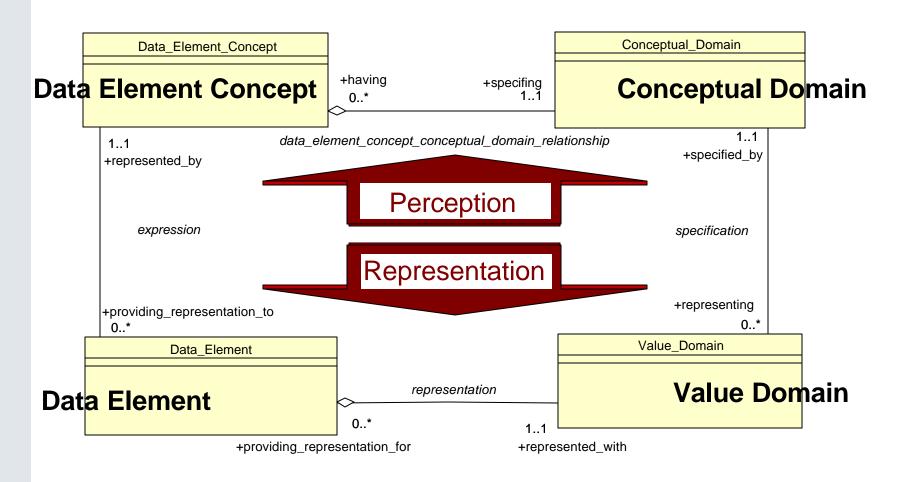
caDSR Background

- Purpose:
 - Support the development and sharing of CDEs that are used as metadata descriptors for NCI-sponsored research
- Goals of caDSR Tools development:
 - Simplify development and creation of ISO/IEC 11179 compliant metadata by CDE Curators
 - Simplify consumption of CDEs by end users and application developers
 - Enhance reuse of CDEs for all
 - Enable semantic consistency across research domains
 - Build linkage to EVS concepts





Basic Metamodel Components

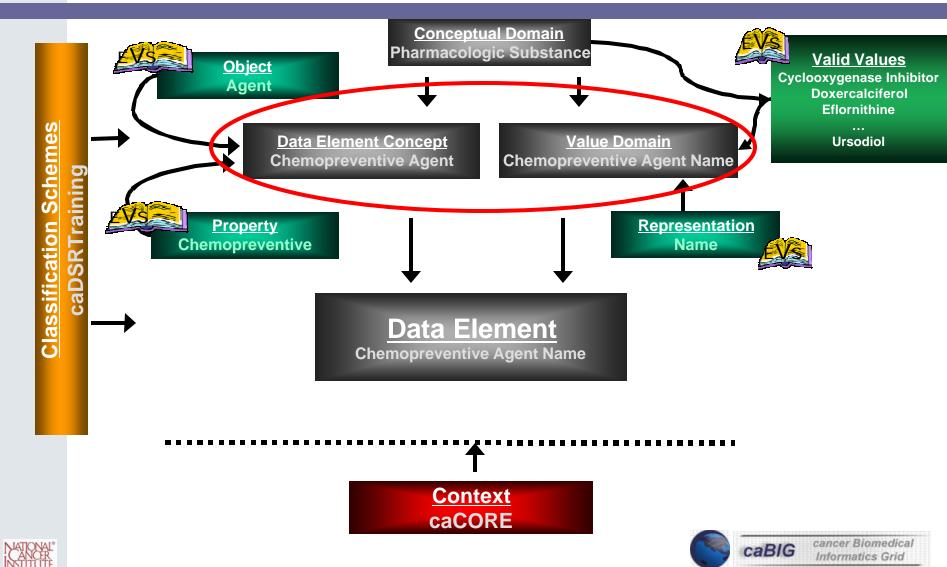








caDSR Implementation ISO/IEC 11179 Model



caDSR End-to-End

But WHY?

So much metadata - So little time

Why bother to link to EVS?

Once I've defined them, how can a CDE be used to optimize research?







caDSR End-to-End

What if?

- You could use the CDEs to gain knowledge?
- You could find other studies or data repositories with data you are interested in analyzing?
- You could leverage EVS to discover hidden relationships between datum?









caDSR and EVS

caDSR is a metadata repository, evolving into a Registry

maintains metadata to permit a user to locate the correct data element defining the characteristics of a piece of datum, an instance of a specific concept, in sufficient detail to be collected and stored on a computer

EVS is a terminology server

rovides services for synonymy, mapping between vocabularies, hierarchical structures, Subconcepts, Superconcepts, Roles, Semantic type, etc.







caDSR End-to-End

Framework:

Protocol → Study, Repository

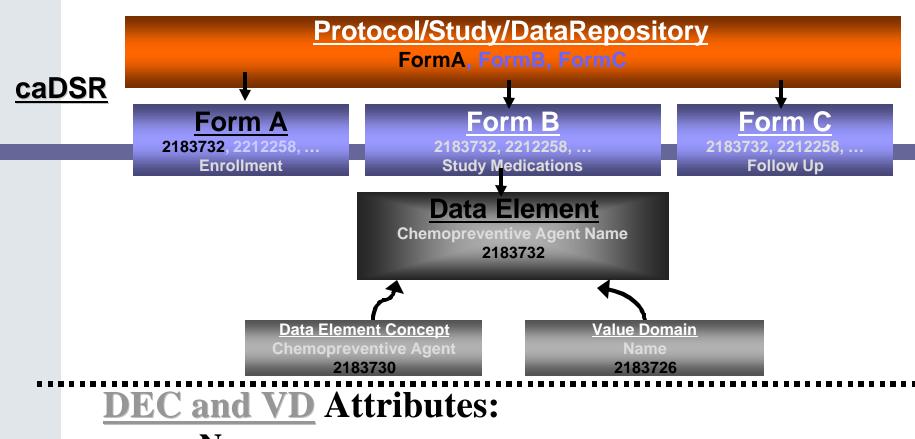
Form → Data collection instrument

Data Element → A specific representation/instance of a concept









Name

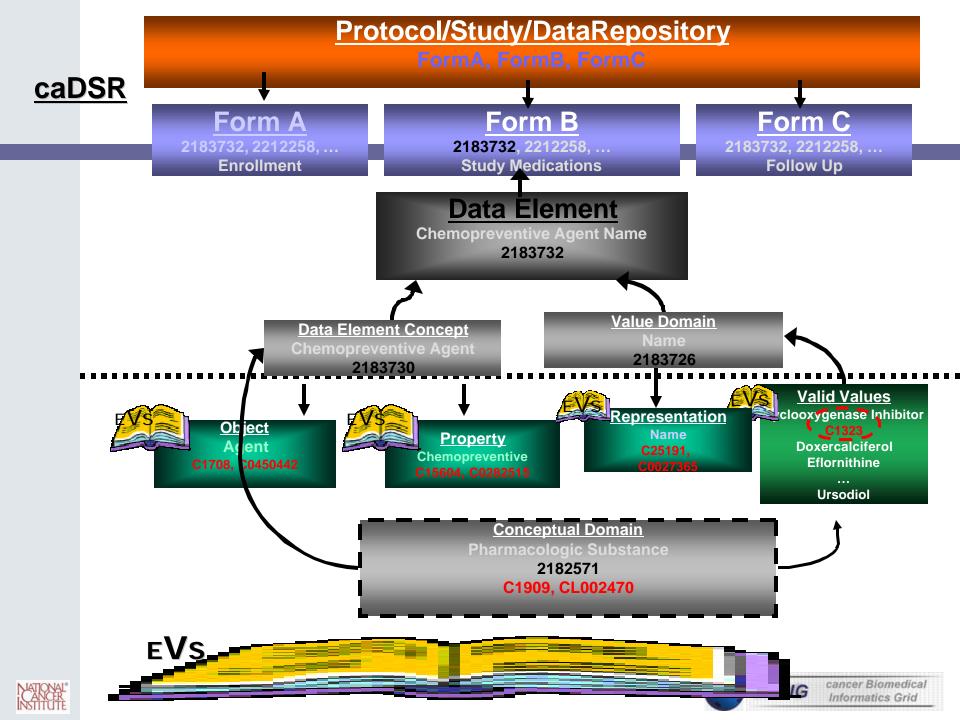
Definition

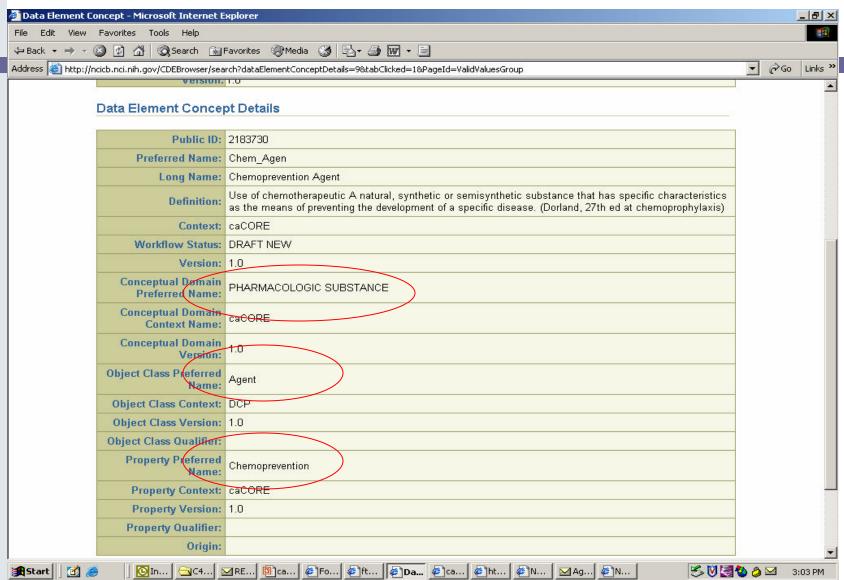
Alternate Name = EVS LINKAGE





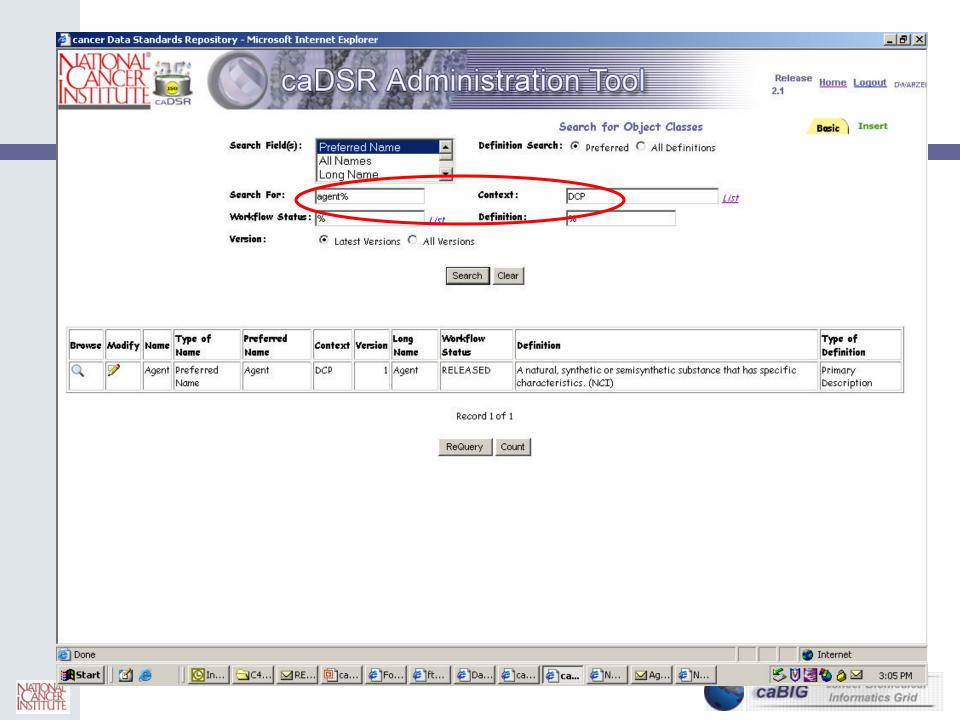


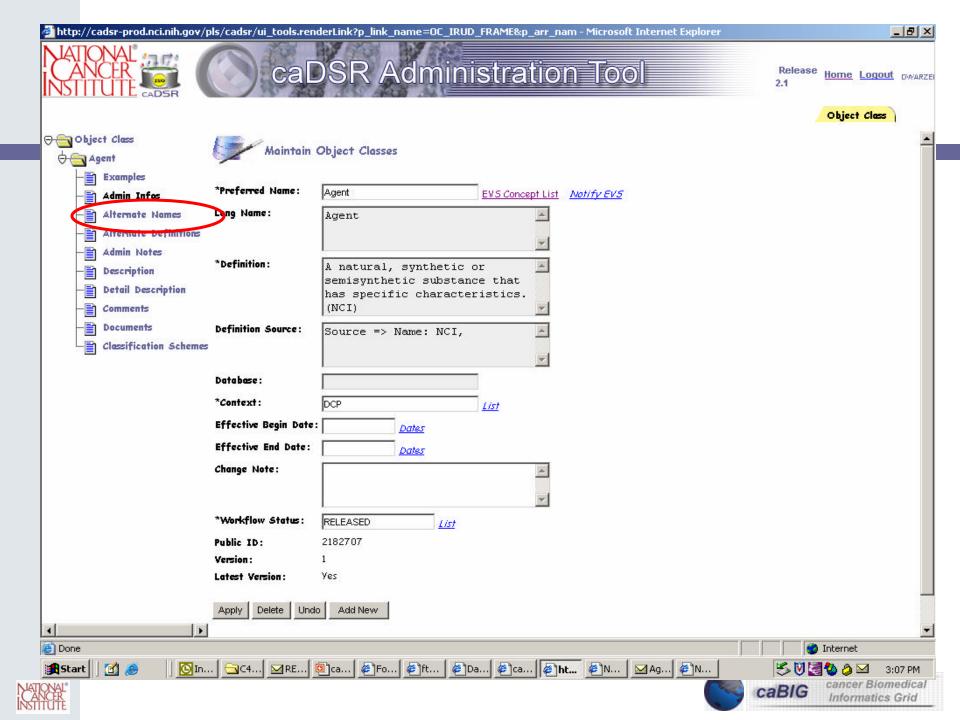


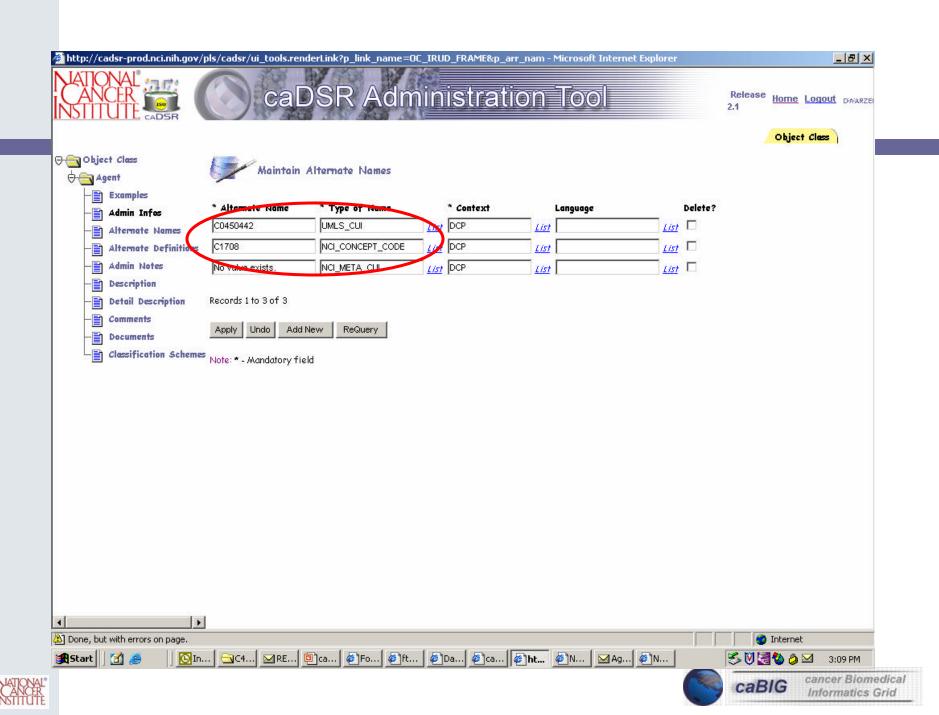


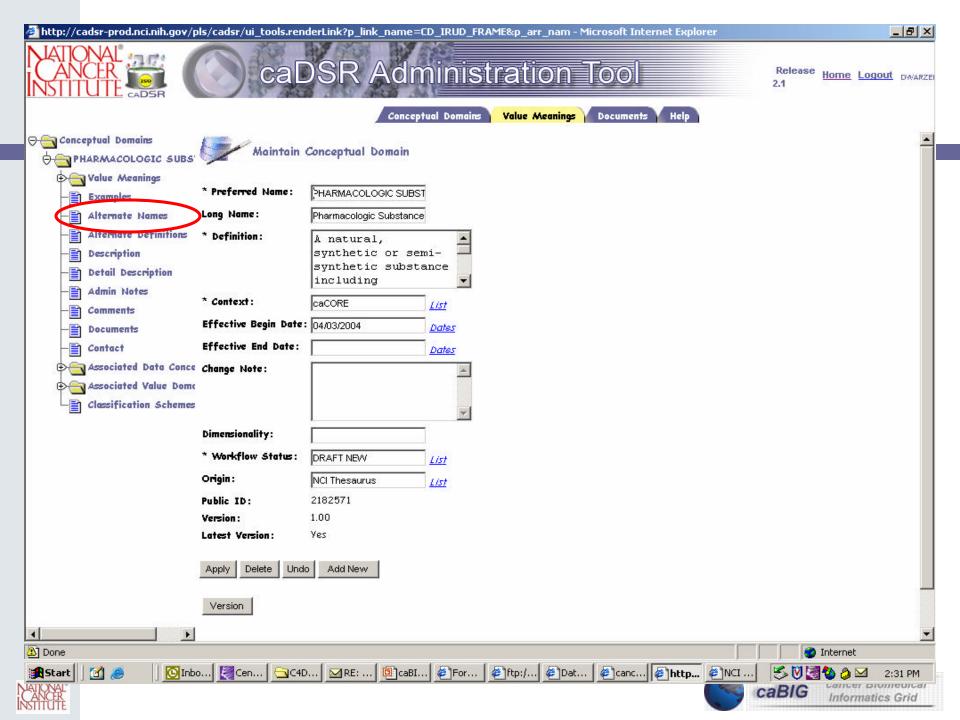


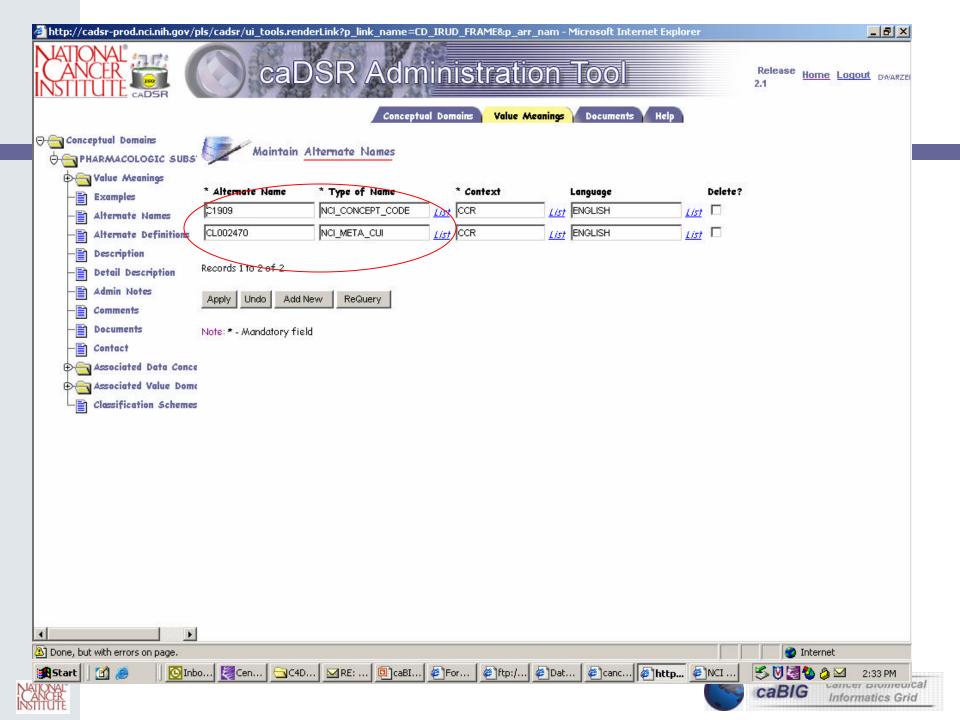








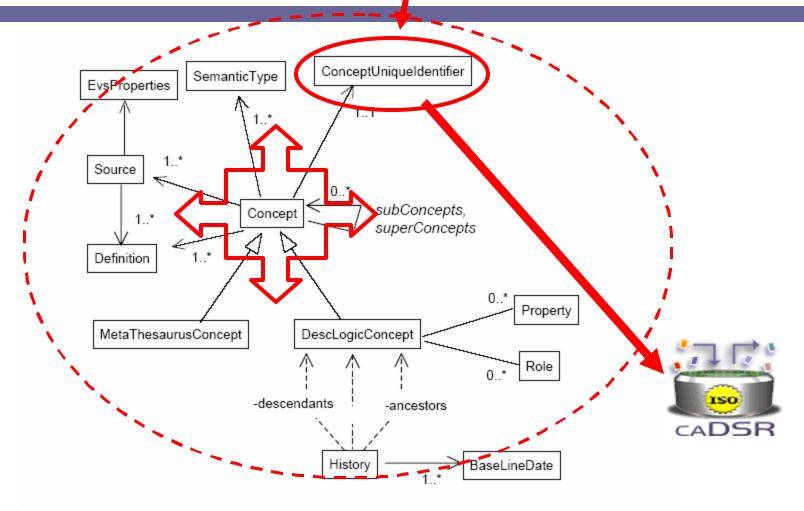




EVS

caDSR Alternate Name EVS Linkage – (Automatically

for Object Class, Property, Rep Term, Valid Value [3.0])

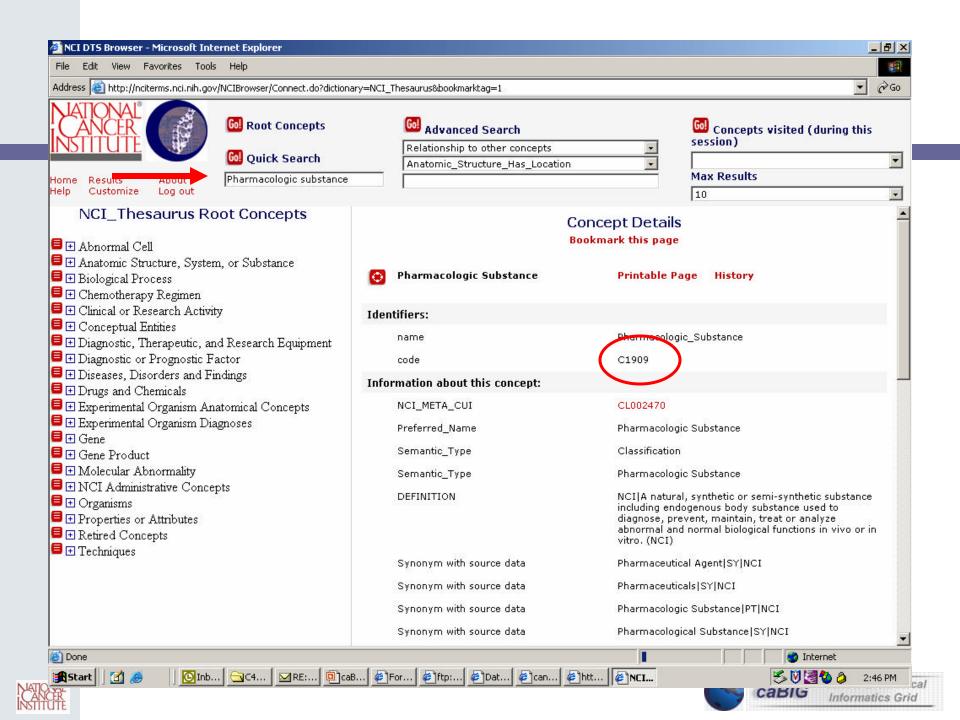


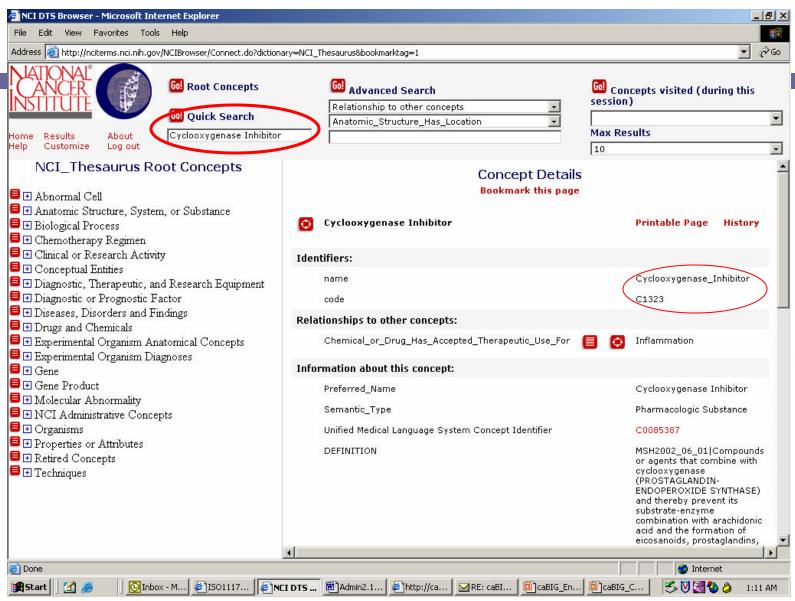






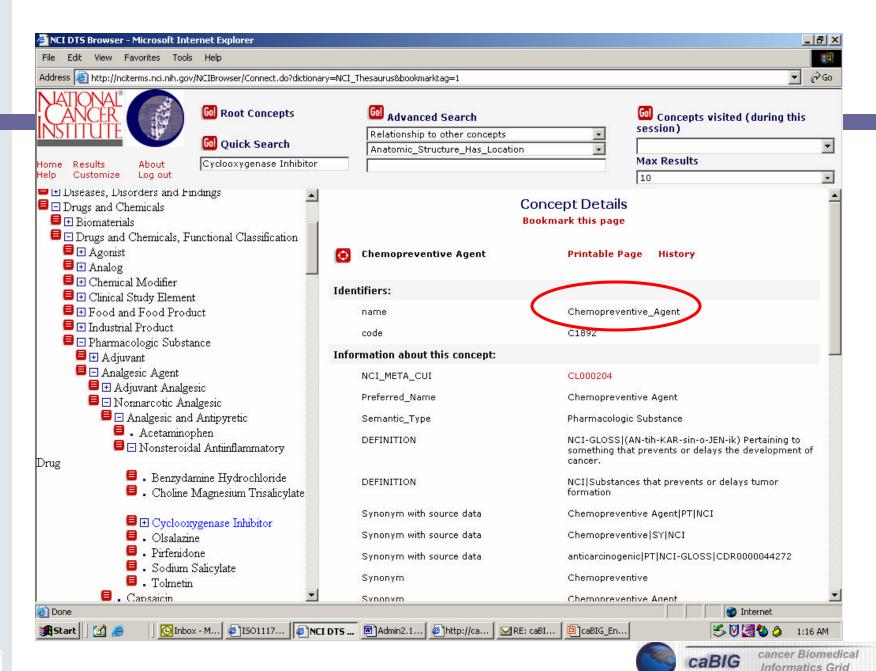




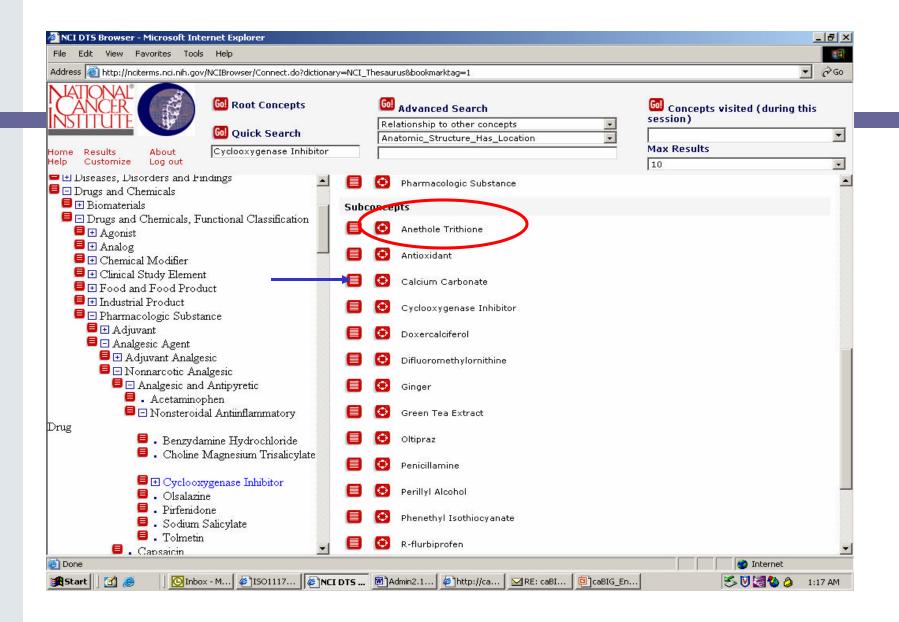






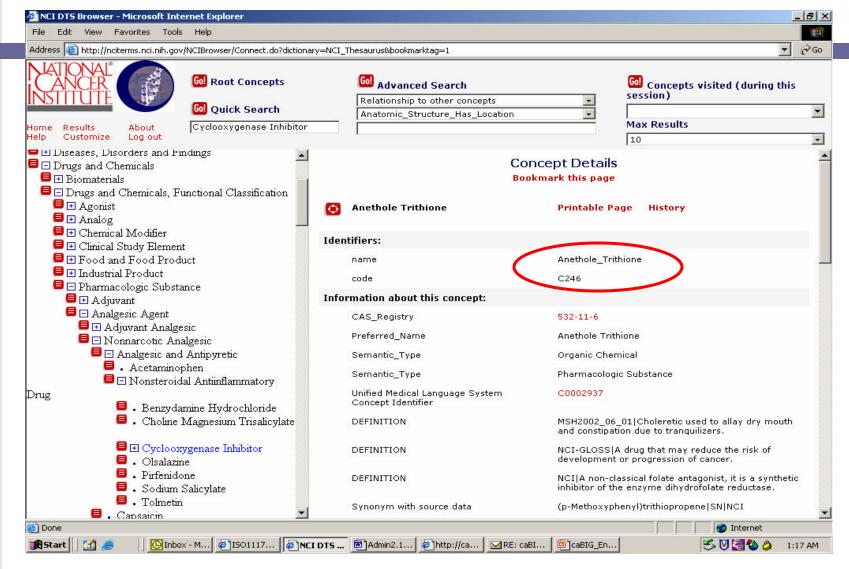










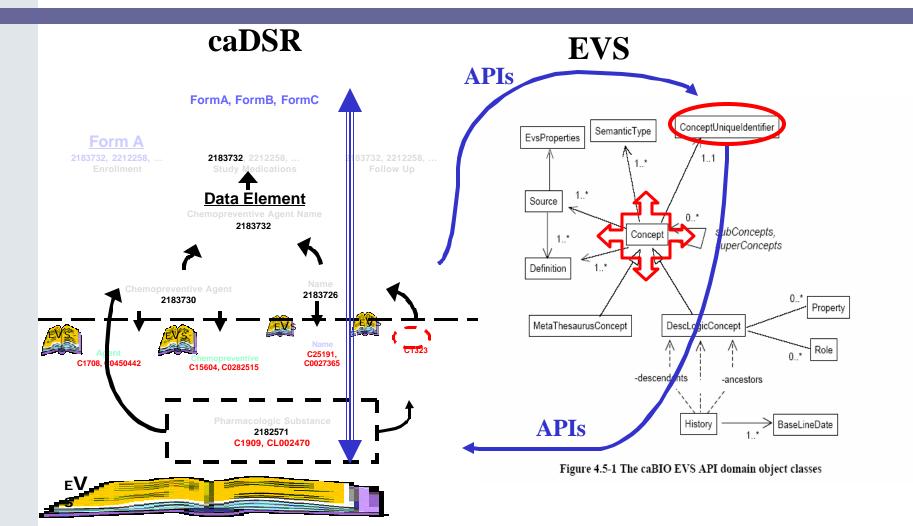








caDSR→ EVS Round Trip









caDSR→ EVS Round Trip

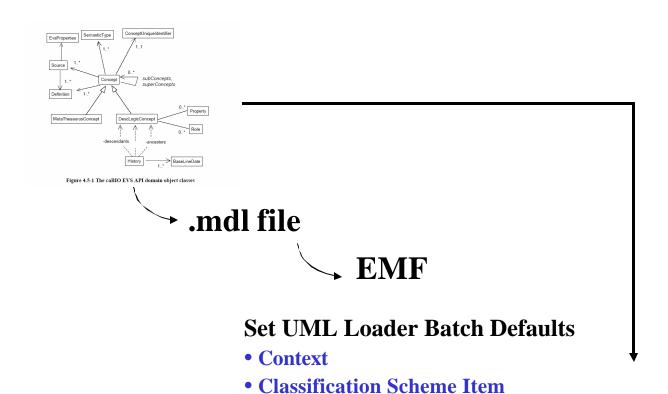
caDSR **EVS** ATTIS ConceptUniqueIdentifie SemanticType What other DEs have this Valid Value? **EvsProperties** What Studies were they collected in? Who is the owner of the protocol? Source Is the resource available for sharing? ıbConcepts. uperConcepts If its UML is loaded is there or can I build an API? How does that compare with my study? Property Which studies collected the related agent? MetaThesaurusConcept icConcept What other CDEs were collected in the study? Role Can I do a meta-analysis? What are the differences in the outcomes? **APIs** BaseLineDate What roles does this agent it play? What other agents play a similar role? (inflammation .54 The caBIO EVS API domain object classes What can Meta tell me about this concept? What are its ingredients?



caBIG cancer Biomedical Informatics Grid



UML Model Loading



Reference Document

Conceptual Domain\







UML Model Transformation

- Create UML in Rational Rose and save it
 - mdl extension
- 2. Access class diagram and elements at UML Loader run-time
 - Eclipse Modeling Framework (EMF)
 - Meta-model and API
- 3. Transform UML to caDSR Metadata
 - Class = Data Element Concept (DEC)
 - Class name = DEC Preferred Name
 - Package qualified Class name = DEC Long Name

- Class Attributes = Data Element
 - Data Element DEC = Attribute's corresponds Class
 - Data Element Long Name = Attribute Name
 - Data Element Preferred Name = Class
 Name + Attribute Name
 - Data Element Value Domain = Generic VD based on java datatype
- Class Associations = Data Element Relationship (DECr)
 - DECr Description = Information such as cardinality, rolename, direction







UML Model Loading Example

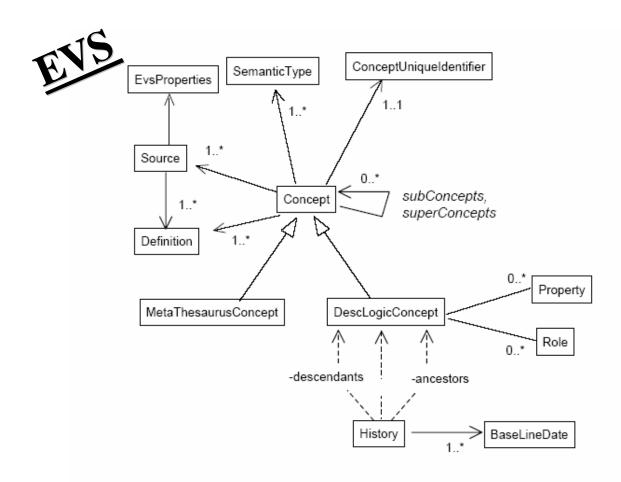


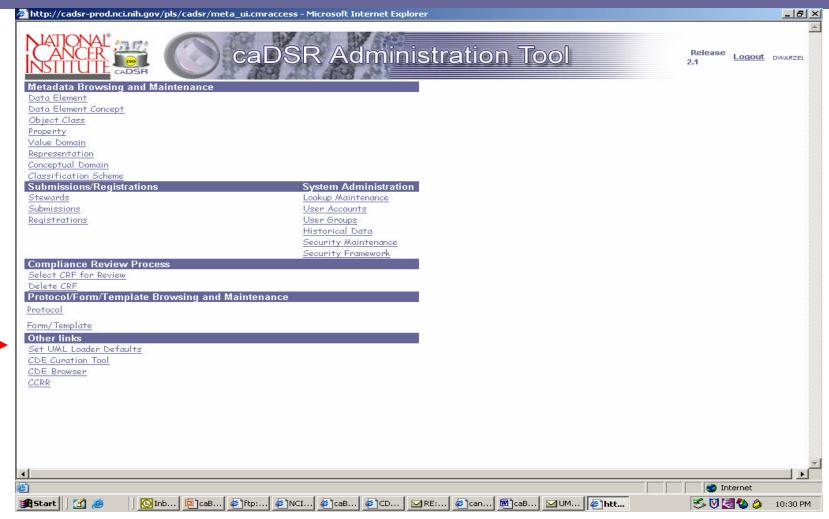
Figure 4.5-1 The caBIO EVS API domain object classes







Set UML Loader Defaults in Admin Tool









Set UML Loader Defaults

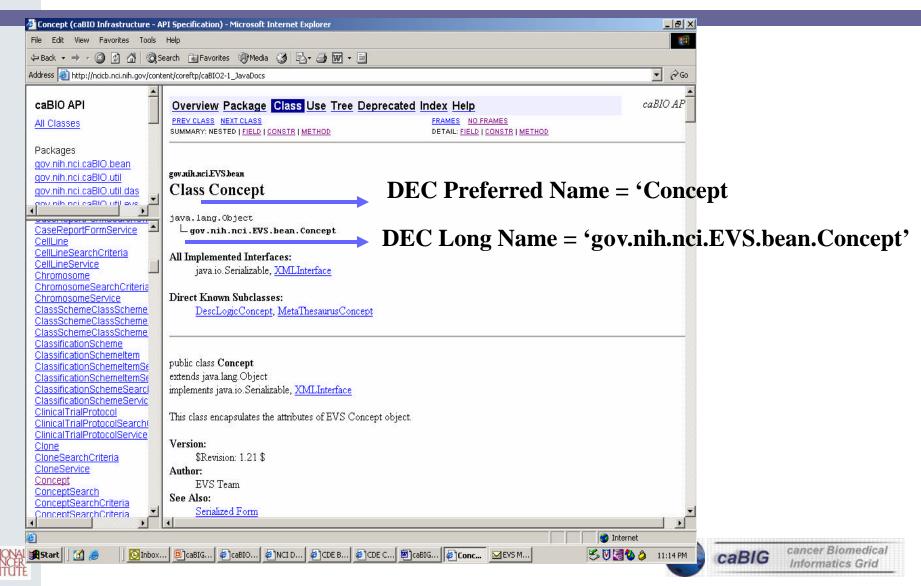
Cancer Data Standards Repose	ry - Microsoft Internet Explorer	_ B ×
NSTITUTE CADSR	caDSR Administration Tool Release Home Logo 2.1	out DWARZE
Load Defaults		
Batch Id:		
Context Name:	caCORE List	
Context Version:	1	
Version :	2.0	
Workflow Status:	DRAFT NEW List	
CD Preferred Name:	BIOINFORMATICS <u>List New</u>	
CD Version:	1	
CD Context Name:	caCORE	
CD Context Version:	1	
VD Preferred Name:	List New	
VD Version:		
VD Context Name:		
VD Context Version:		
Reference Document Name:	EVS-2.1.mdl(May2004)	
Reference Document Type:	REFERENCE List	
Reference Document Text:	EVS Domain Object Model	
Reference Document Name URL:		
Classification Scheme:	JAVA PACKAGES <u>List</u>	
Class scheme Item:	EVS	
Apply Clear		
@ Done	internet	
Start 6	n	10:36 PM





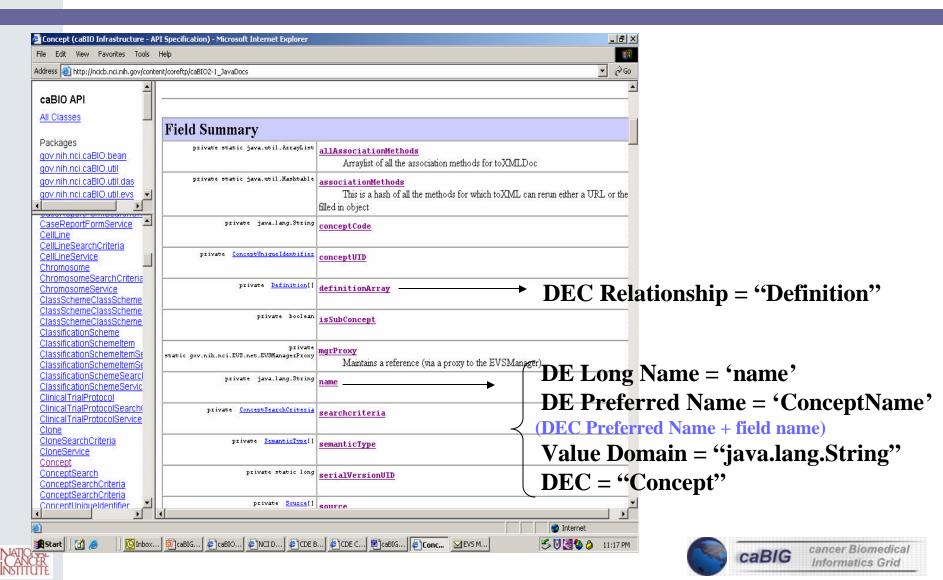


UML Model Transformation



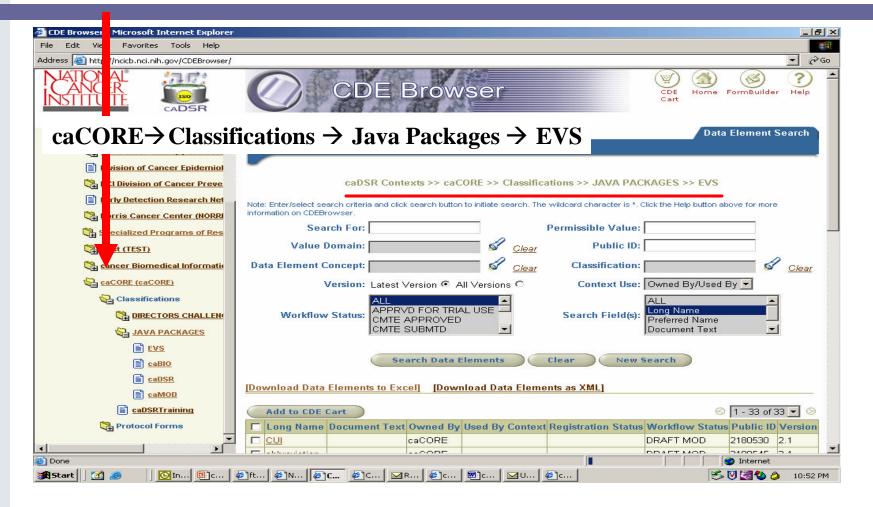


UML Model Transformation





UML Models In CDE Browser

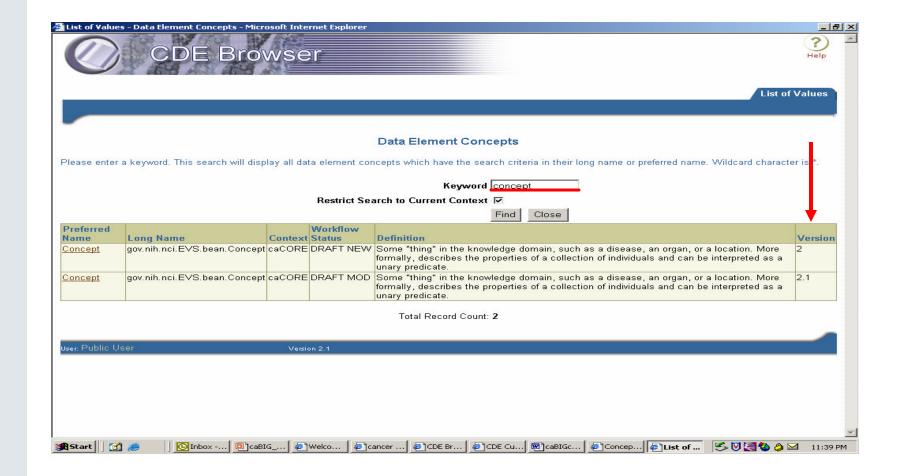








UML Models In CDE Browser

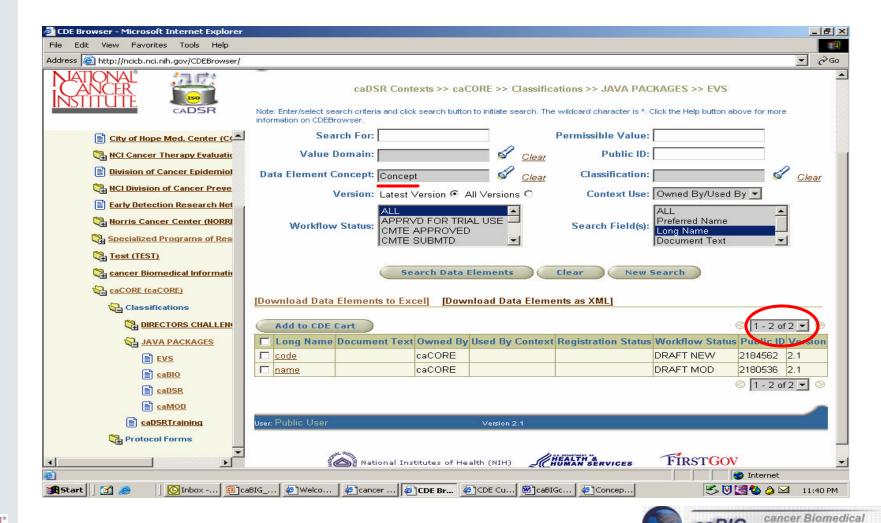






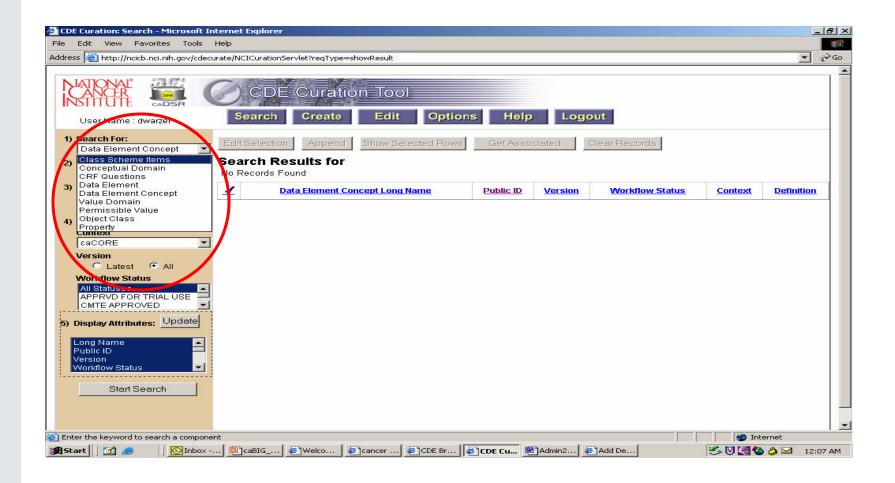


UML Models In CDE Browser

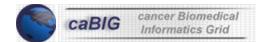


Informatics Grid



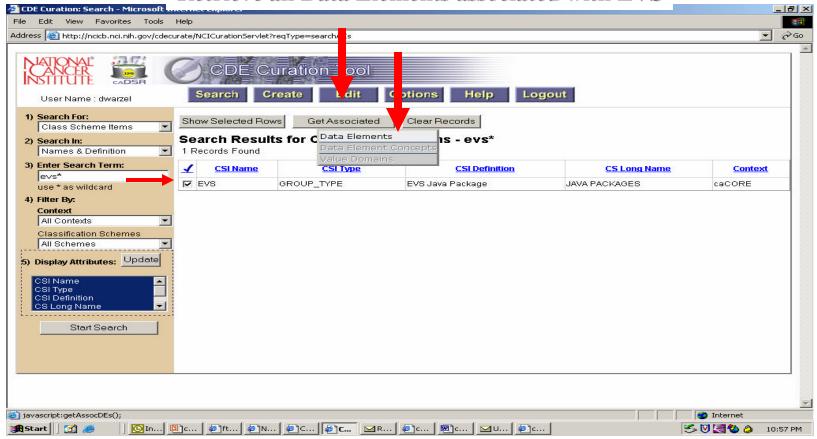








Retrieve all Data Elements associated with EVS

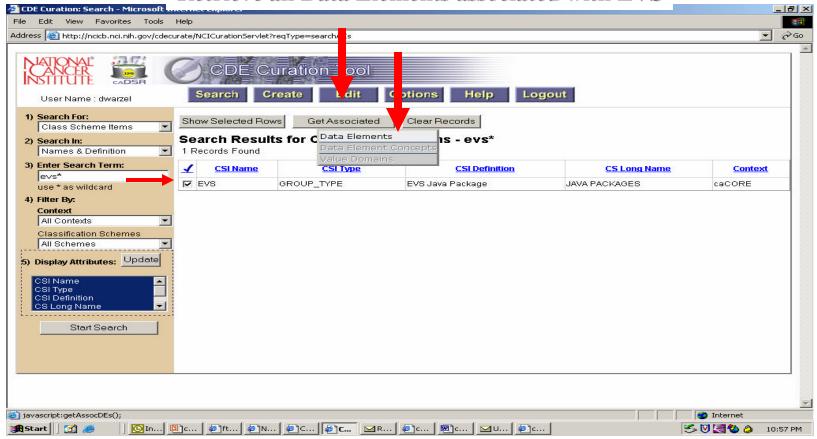








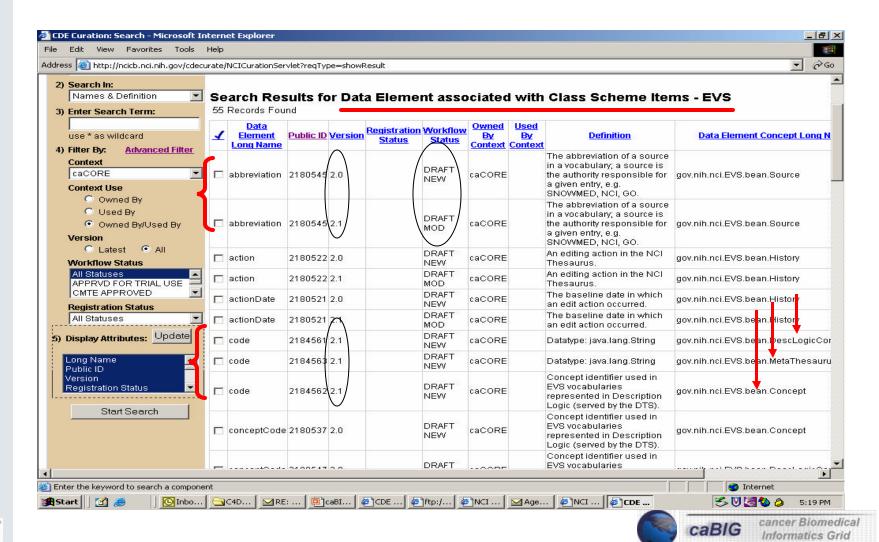
Retrieve all Data Elements associated with EVS







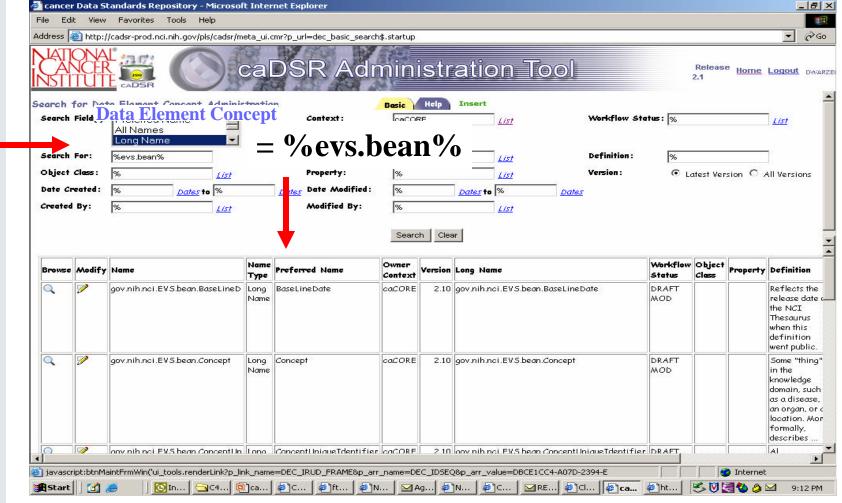








UML Model in Admin Tool

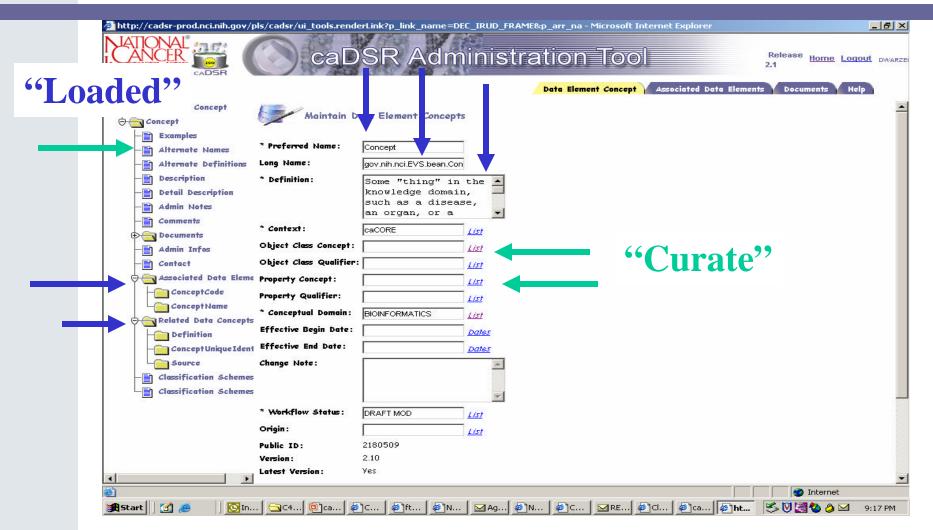








UML Model Loading Example



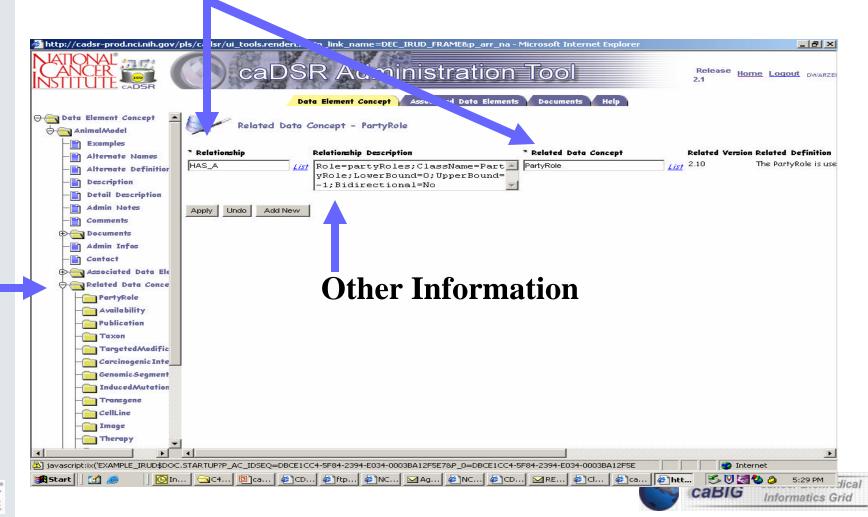






UML Model Loading Example

Class Associations

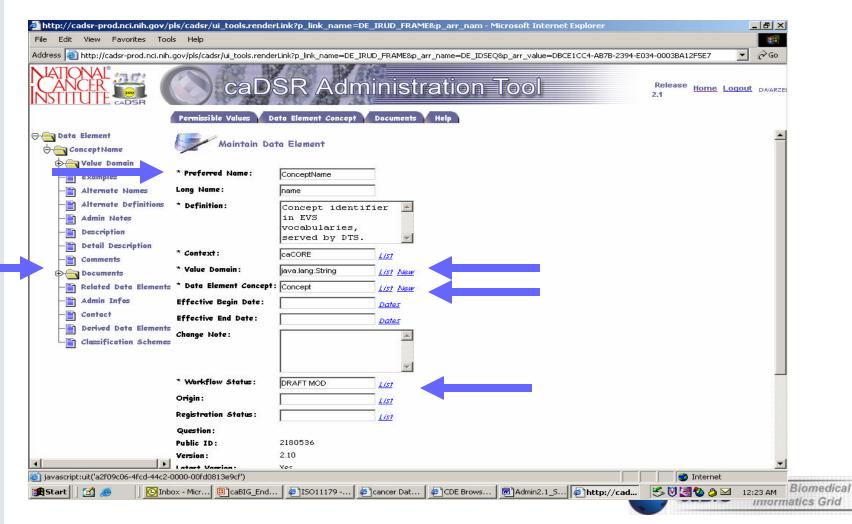






UML Model Loading Example

Class Attributes





Operating Environments

- Database Repository
 - Oracle 8i
- Administration Tool
 - Oracle PL/SQL, Oracle 9i Application Server
- CDE Browser
 - Java, Oracle 9i Application Server
- CDE Curation Tool
 - Jakarta Tomcat
- Case Report Form (CRF) Loader
 - Visual Basic, MS Excel GUI, Oracle 8i Client
- Compliance Review Tool (CRT)
 - Visual Basic, MS Excel GUI, Oracle 8i Client
- CDE Compliance Review Response Tool (CCRR)
 - Java, JSP, Java Beans, Oracle 9i Application Server





caDSR Team

NCI

- Peter Covitz
- Denise Warzel
- caDSR Context Administrators

ScenPro

- Bill McCurry
- Mike Gately
- Ken Sims
- Mark Swenholt
- Jennifer Brush
- Tom Phillips
- Mike Holck
- Smita Hastak
- Wendy Verhoef

Oracle

- Edmond Mulaire
- Ram Chilukuri
- Prerna Aggarwal
- Christophe Ludet
- Daniel Ladino
- Science Applications International Corporation (SAIC)
 - Kathleen Gundry
 - Tommie Curtis
 - Brenda Maeske
 - Mead Walker





Questions?



- caDSR Home Page
 - http://ncicb.nci.nih.gov/core/caDSR
 - Biweekly Software Meetings Monday's 3:00 pm EST (Warzel)
 - Alternate bi-weekly Content Meetings 3:00 pm EST (Curtis)



- Administration Tool
- http://cadsr-prod.nci.nih.gov
 - CDE Curation Tool
 - http://ncicb.nci.nih.gov/cdecurate

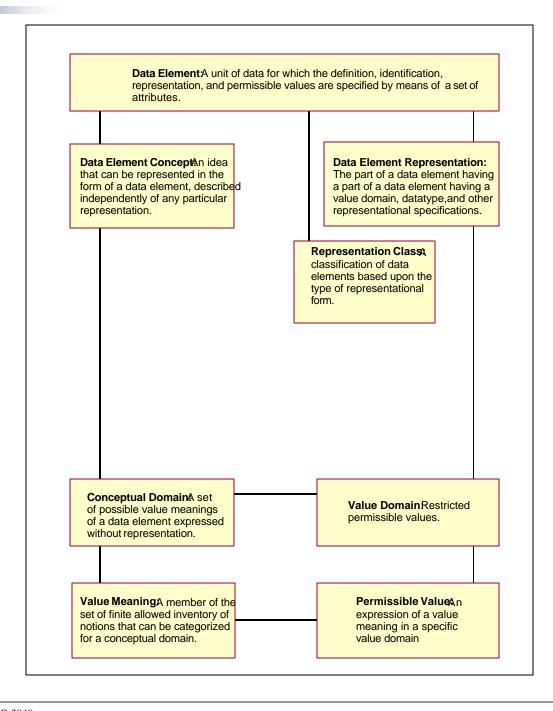


- CDE Browser
 - http://ncicb.nci.nih.gov/cdebrowse

- warzeld@mail.nih.gov
- **303-722-9446**







Terms and Definitions for ISO/IEC 11179-3